Department of
ELECTRICAL ENGINEERING

UNIVERSITY OF NOTRE DAME, NOTRE DAME, INDIANA

Final Report:
Sub-Band/Transform Compression
of Video Sequences

Grant No. NAG3-1186
January 11, 1991 to January 10, 1992

Principal Investigator:
Prof. Ken Sauer

Co-Principal Investigator:
Prof. Peter Bauer

Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Contents

1 Introduction

This report summarizes the progress on compression of video sequences during the second period of grant no. NAG3-1186. The overall goal of the research was the development of data compression algorithms for high-definition television (HDTV) sequences, but most of our research is general enough to be applicable to much more general problems. As the title suggests, we have concentrated on coding algorithms based on both sub-band and transform approaches.

Two very fundamental issues arise in designing a sub-band coder. First, the form of the signal decomposition must be chosen to yield band-pass images with characteristics favorable to efficient coding. The idealized decomposition is compromised by the width of transition bands of the band-splitting filters, which should be of practical design, with as few coefficients as possible. Thus, the design of band-pass filters occupies an important place in our research. A second basic consideration, whether coding is to be done in two or three dimensions, is the form of the coders to be applied to each sub-band. Here again, computational simplicity is of essence.

We have investigated both of the above issues during the past year, and will discuss them in the following sections. In Section 2 we will review the first portion of the year, during which we improved and extended some of the previous grant period's results. Section 3 will cover the pyramid nonrectangular sub-band coder limited to intra-frame application. Perhaps the most critical component of the sub-band structure is the design of bandsplitting filters. We apply very simple recursive filters, which operate at alternating levels on rectangularly sampled, and quincunx sampled images. We will also cover the techniques we have studied for the coding of the resulting bandpass signals. In Section 4, we discuss adaptive three-dimensional coding which takes advantage of the detection algorithm developed last year.

To this point, all the work on this project has been done without the benefit of motion compensation (MC). Motion compensation is included in many proposed codecs, but adds significant computational burden and hardware expense. We have sought to find a lower-cost alternative featuring a simple adaptation to motion in the form of the codec. In sequences of high spatial detail and zooming or panning, it appears that MC will likely be necessary for the proposed quality and bit rates.

Three graduate students and one post-doctoral researcher have benefitted from the finan-

cial support of Grant NAG3-1186. During the first three months of the year, Coleen Jones finished and defended her thesis on sample allocation through detection of dynamics and exploitation of properties of the human visual system. She is now employed by the National Institute of Standards and Technology in Boulder, Colorado, doing research on quantitative standards of image quality. Qian Wei moved on to a new appointment at the University of South Florida after helping us with nonlinear filter design continued from the previous grant period. Filter design is currently the pursuit of Jie Wang, a second-year Ph.D. student. Xiaohui Li replaced Ms. Jones last spring, and has done most of the work on design of subband coders, and implementation of others, such as the JPEG standard still-image coder.

2   Completion of Previous Topics

For the first months of 1991, after the completion of the previous grant period, Coleen Jones and Qian Wei remained with the project, and extended their work from the previous final report's information. Mr. Wei's work during the later period was in more extensive simulation of our proposed nonlinear preprocessing filter[1]. His results established more dependable guidelines for choosing the IIR filter coefficient in the linear module of the preprocessing filter.

At the end of the previous period, we faced the problem of estimating the partition constant used to normalize the Gibbs distribution describing the pixel level MRF in our model for binary fields of detected active pixels. The general form of the density function with external field parameter γ_1 and inter-pixel bond β is

$$p(\mathbf{h}) = \frac{\exp\left\{-\sum_{j \in \mathcal{B}_i} \left[\beta \sum_{k \in \mathcal{N}_j} |h_j - h_k| + \gamma_1 h_j\right]\right\}}{Z(\beta, \gamma_1)},$$

where the h_i are the binary results of pixel-level detection of temporal change. The constant is dependent on the pixel-level detection probabilities, which are periodically updated after noise estimation. Our goal was to approximate the denominator constant with a form which would allow a simple update, and recomputation of threshold value for the maximum-likelihood block classification. Our solution, which appears to work well, was approximating the constant as the *product* of two terms, representing the interpixel cost, and the external field:

$$\tilde{Z}(\beta, \gamma_1) = C(\beta)(1 + \exp(-\gamma_1))^N = C(\beta)(1 - p_1)^{-N},$$

with p_1 the probability of each pixel's being detected as active. This form is similar to the zero connectivity case, and equal to it when $\beta = 0$, implying $C(\beta) = 1$. Since β is constant through the sequence, this number need only be chosen or estimated once, after which the constant, and therefore the block threshold can be simply adjusted. Ms. Jones also performed simulations on other sequences, assuring the more general applicability of her work. Further details are available in [2, 3, 4].

3   Pyramid Coding with Non-Rectangular Sub-Bands

Sub-band coding methods have been shown useful and flexible for many applications in speech and image processing[5, 6]. Their performance in image compression is comparable

to that of the ubiquitous discrete cosine transform (DCT), but error is localized in frequency, rather than space, and effects such as block boundary visibilities are naturally avoided. The usual frequency-domain geometry of decomposition is a series of rectangles. This allows the use of separable filters with design in one dimension. In the following discussion, we will elucidate only the algorithms for coding the luminance portion of the video sequence. The lower bandwidth chrominance images are compressed at lower resolution by a JPEG transform-based technique[7].

As was demonstrated in our previous work[2] and that of other researchers, the human visual system (HVS) has curves of constant sensitivity in the spatial frequency domain which are approximately diamond-shaped. Thus lowpass filtering for data reduction is ideally done using non-rectangular passbands. Ansari, et al.[8] exploited this fact in a sub-band coder which lowpass filtered each image, followed by a two sub-band decomposition of the remaining signal shown in Fig. 1. The diamond-shaped passband allows a quincunx downsampling of the original image with an immediate reduction of sampling density by a factor of two. The quincunx image in [8] underwent one further band separation, with the lowest band coded by the DCT.

If high resolution imagery is to be available to devices with widely varying sampling rates, transmitted through channels of varying capacities, it may be necessary to bandlimit a given image at several levels. Since the diamond-shaped sensitivity contours hold over a broad range of spatial frequencies, it appears desirable to have available a hierarchy, or pyramid, of frequency sub-bands. This is also true for temporally adaptive three-dimensional coding of image sequences, as illustrated in [9]. The form of the proposed decomposition is simply a generalization of that in Figure 1, with the inner square further decomposed recursively. While the pyramid need not stop at a particular level, we have found a five-band decomposition adequate for our purposes.

The decomposition process can be understood by considering the transition from one level of the pyramid to one other. For simplicity, we describe the first, which is illustrated in Fig. 2. By bandlimiting the image to the diamond shape, we can produce two images whose sum is the original. Each of the resulting images is now sampled at a rate twice the necessary density. Provided aliasing is excluded, each image can be decimated by a factor of two in a quincunx pattern, as shown in Fig. 3. Should we wish to recombine to recover the original image, each sub-band can be interpolated to the original sampling rate, and
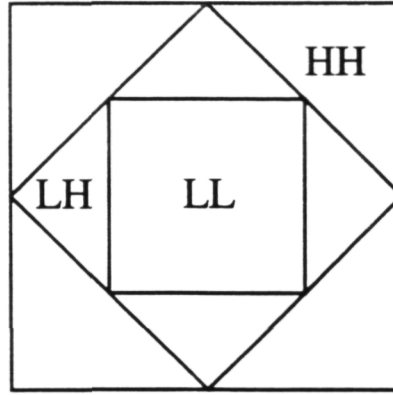
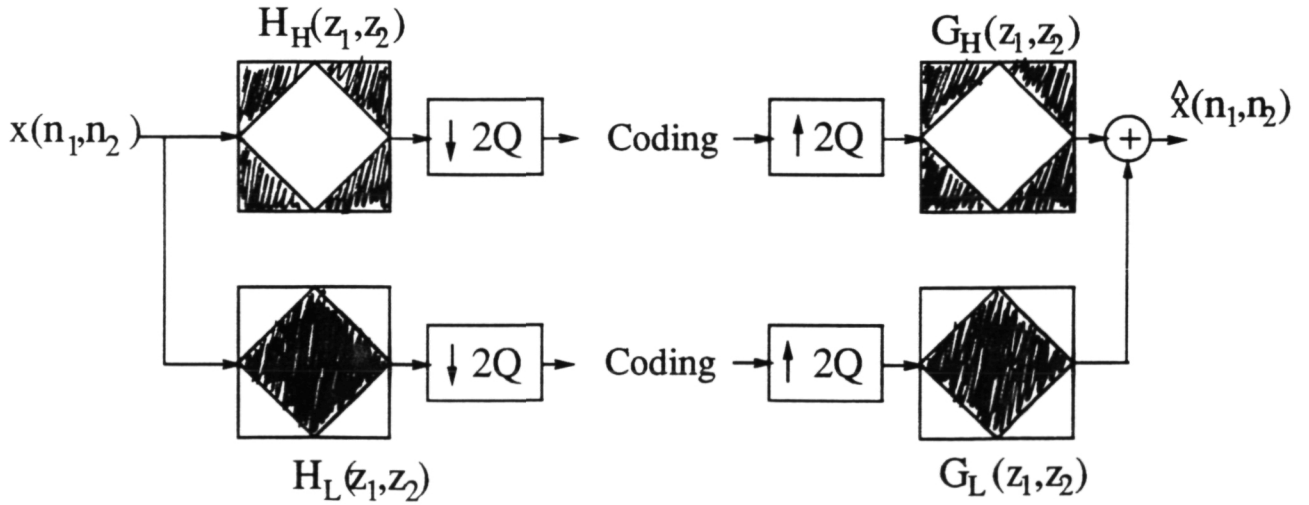Figure 1: Shape of sub-band decomposition proposed by Ansari, et al.

Figure 2: Single stage of diamond-shaped sub-band decomposition. "2Q" indicates a quincunx decimation reducing rate by 2. Shaded portions represent the passbands of corresponding band-splitting filters. Channel effects are not included here.
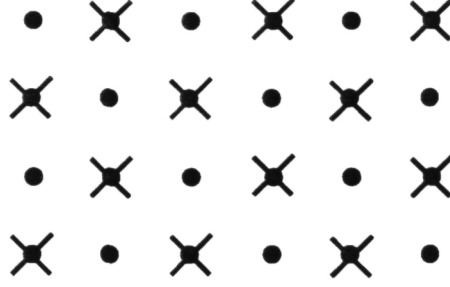
Figure 3:   Quincunx sampling pattern resulting from "2Q" decimation after separation of previous figure.

bandpass filtered before summing.

Even when some aliasing exists in the decimated images, perfect reconstruction filter banks can produce aliasing cancellation in the final result to erase error[10]. In the absence of coding and channel errors, the output in Fig. 2 can be written as

$$\hat{X}(\mathbf{z}) = [H_L(\mathbf{z})G_L(\mathbf{z}) + H_H(\mathbf{z})G_H(\mathbf{z})]X(\mathbf{z}) + [H_L(-\mathbf{z})G_L(\mathbf{z}) + H_H(-\mathbf{z})G_H(\mathbf{z})]X(-\mathbf{z}). \quad (1)$$

Here we have adopted the vector notation $\mathbf{z}$ for (z_1, z_2). Provided the second term on the right-hand side is zero, aliasing will be cancelled at reconstruction. Perfect reconstruction follows also if the first term equals 1.

Note that the two quincunx-sampled sub-band images may be thought of as being represented on a rectangular grid, rotated by 45 degrees from the original. In this orientation, we can think of again lowpass filtering the low-frequency image to separate bands as at the first level. The resulting decimated lowpass image will be sampled on a rectangular grid which is decimated by a factor of two in each coordinate relative to the original. This process can be continued in a pyramid fashion, with successive lowpass images of smaller size, alternating between rectangular and quincunx grids.

We have studied the filter design and coding aspects of the pyramidal non-rectangular sub-band coder, and implemented a version which appears to perform well at intraframe bit rates of about 1.0 to 1.5 bits/pixel for color imagery. These rates would allow transmission of HDTV sequences at about 30-45 Megabits/sec with very limited hardware expense.

3.1 Design of Recursive Bandsplitting Filters

Recursive (IIR) filters for sub-band coding offer good performance with significantly reduced computational cost[11, 12]. We use only IIR bandsplitting filters in the work presented here. Two fundamental filter modules form the core of the entire filter structure for the tree-structured sub-band decomposition. For the separation of bands at the coder (analysis filters), we employ the filters described by:

$$H_L^{(1)}(z_1, z_2) = \frac{1}{2}[1 + z_1 T(z_1 z_2) T(z_1 z_2^{-1})] \quad \text{(Lowpass)} \tag{2}$$

$$H_H^{(1)}(z_1, z_2) = \frac{1}{2}[1 - z_1 T(z_1 z_2) T(z_1 z_2^{-1})] \quad \text{(Highpass)} \tag{3}$$

$$\text{where } T(z) = \frac{az + 1}{z + a} \tag{4}$$

$T(z_1 z_2)$ and $T(z_1 z_2^{-1})$ are one-dimensional all-pass filters, operating on an image in diagonal directions. The desired frequency response is created by phase cancellations. Higher order filters substituted for $T(z)$ have narrower transition bands, but do not necessarily yield better results. These second-order prototypes guarantee perfect reconstruction if the synthesis filters are chosen to be causality-inverted, scaled versions of the analysis filters[8]. That is, the conditions imposed above by (1) are satisfied by

$$G_L(z_1, z_2) = 2H_L(z_1^{-1}, z_2^{-1}), \tag{5}$$

$$G_H(z_1, z_2) = 2H_H(z_1^{-1}, z_2^{-1}). \tag{6}$$

The most effective implementation of the filter blocks H_L, H_H can be achieved using the concatenation in Fig. 4.

Note from (2)-(4) that a single frequency response, either that of the lowpass or highpass module, specifies the entire system. The frequency responses for typical implementations of $H_L(z_1, z_2)$ are given in Fig. 5 for the first-order case. Though these filters clearly allow aliasing, they possess aliasing cancellation properties common in perfect reconstruction filter banks[10], and will allow error-free reconstruction of the image at the receiver. The response here depends on a single parameter, a, which yielded best results with $0.25 < a < 0.35$. We have observed the best performance in the n-th order case with $a_1 = a_2 = \ldots = a_n$. The implementational and stability considerations for these filters are summarized in the Appendix.
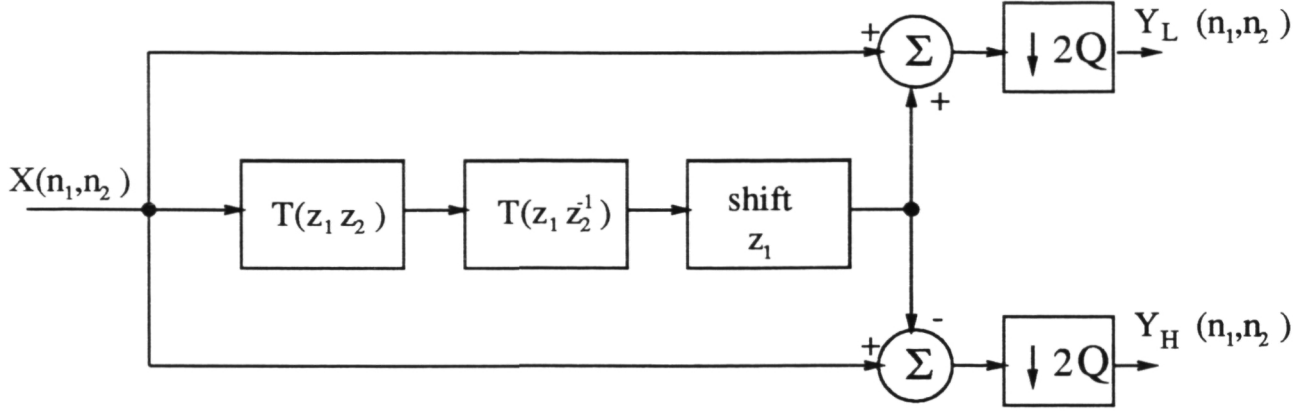
Figure 4: Single stage of nonrectangular analysis filter tree.

The same filters are used at all other levels of the pyramid:

$$H_L^{(\nu+1)}(z_1, z_2) = H_L^{(\nu)}(z_1 z_2, z_1 z_2^{-1}) \tag{7}$$

$$H_H^{(\nu+1)}(z_1, z_2) = H_H^{(\nu)}(z_1 z_2, z_1 z_2^{-1}) \tag{8}$$

where $H_L^{(\nu)}$, denotes the filter transfer function at level ν in the analysis/synthesis structure. The filters at any level can operate directly on the image resulting from the previous level, after the image is simply rotated by 45° and subsampled. With this structure, we can use the same filter description at each level of the pyramid, simplifying both the design process and implementation. Fig. 6 is the frequency response of the lowpass section of the second level of the pyramid. Note the approximate square passband centered at the origin in frequency.

A sketch of a portion of the complete pyramid is given in Fig. 7. In theory, the decomposition may be carried out until the image consists of a single pixel. In terms of entropy considerations, the goal of sub-band decomposition is separation of the image into bandpass functions which have much more flat spectral content than the original. If the power spectral density within the passband of a given image is flat, the fully decimated image may be coded most efficiently by a PCM coder, since samples will be uncorrelated. For practical coding purposes, a small number of levels in the pyramid appears adequate.

The pyramid sub-band separation is closely related to wavelet decomposition[13]. The greater bandwidth of the higher frequency sub-bands yields greater spatial resolution, while the lower frequencies, which concentrate at the top of the pyramid, are highly resolved in frequency, while having low spatial resolution at original sampling rates. While wavelets have received a good deal of attention of late, no clear practical advantages of the wavelet view of image compression has yet been demonstrated.
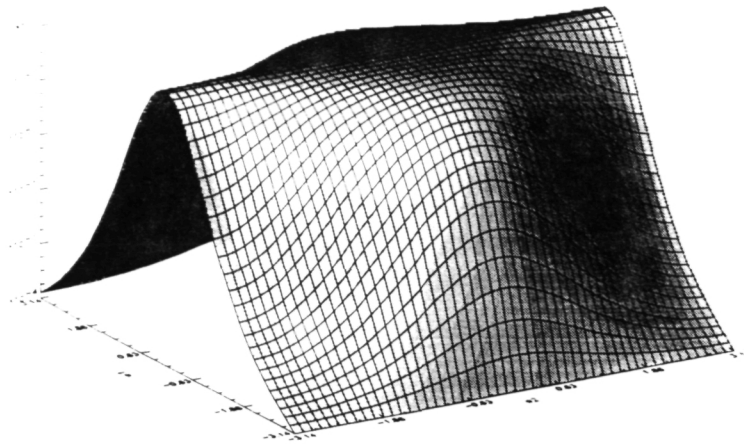
(a)
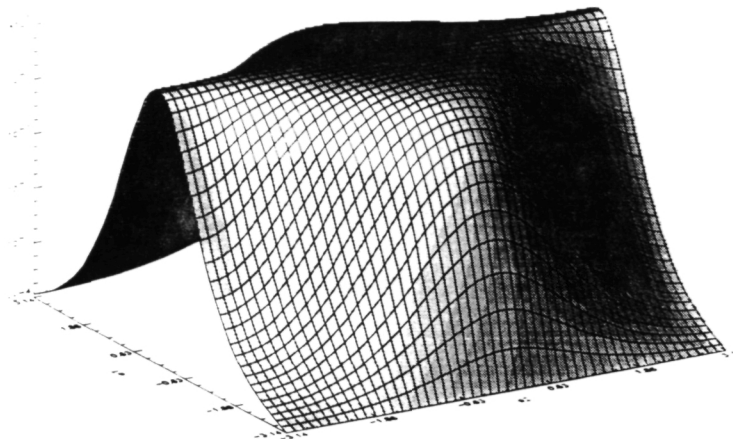
(b)
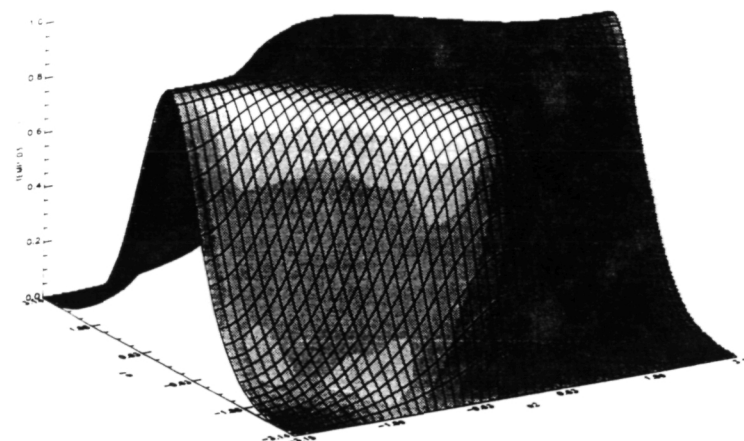
(c)

Figure 5: Frequency response of recursive, separable lowpass filtering operation $H_L(z_1, z_2)$ for three values of the coefficient. (a) $a = 0.2$; (b) $a = 0.25$; (c) $a = 0.33$.
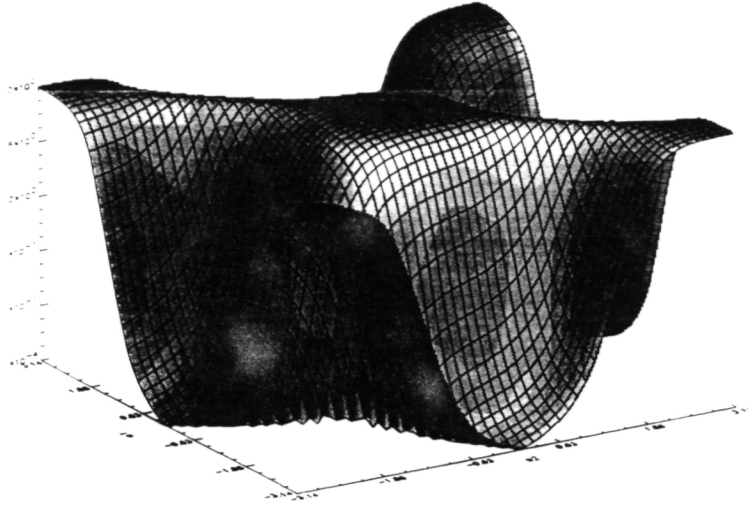
9

Figure 6: Frequency response of lowpass section of decomposition at second level of pyramid. This filter operates on the rotated, quincunx-sampled image, filtering horizontally and vertically in order, followed by the addition and subtraction expressed in (2) and (3).
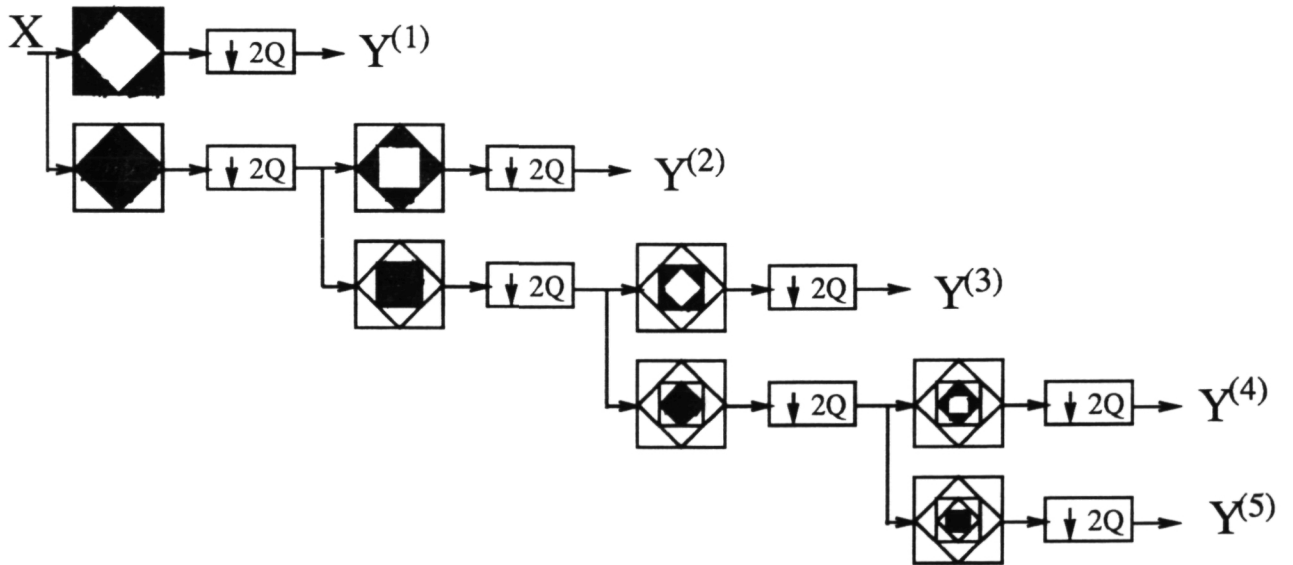

Figure 7: Analysis side of nonrectangular sub-band pyramid, producing signals $\{Y^i\}$. Superscripts correspond to ν in (7) and (8). Synthesis side reconstructs the image in the inverse order, interpolating in the quincunx pattern at each step between levels.
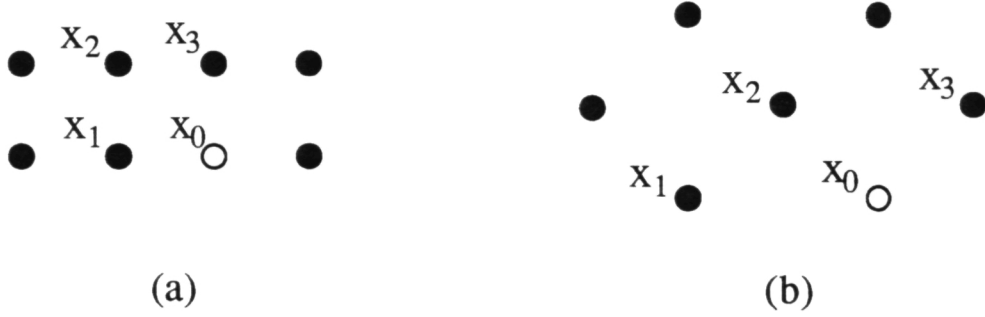
<div align="center">(a)　　　　　　　　　　　　(b)</div>

Figure 8: (a)Causal prediction pattern for rectangularly sampled sub-band images. The scan proceeds from left to right, and top to bottom. (b) Similar pattern for quincunx-sampled images. This pattern results from our maintaining original actual directions of scan for software simplicity. By rotating scan into a rectangular pattern, one could use the same form predictor in (b) as (a).

3.2 Coding of Nonrectangular Sub-Bands

As mentioned above, if sub-bands are separated to the point where the spectrum is flat within each, we may simply elect to code each sample in each decimated sub-band image individually. In practice, we seldom work with such cases. Many types of coders have been employed on sub-bands, including DCT[8], DPCM[5], and vector quantization[14]. DCT-based methods may be useful in the lowest frequency band, but some of the characteristics which make it desirable as an unaccompanied technique are lost. Most DCT methods, such as the JPEG standard, gain a large portion of their compression by taking advantage of large numbers of coefficients quantized to zero. In a pyramid of more than three levels, the low band, which may be decimated in a 4×4 pattern, has far fewer of these zeros. Considering purely compression ratios in intraframe coding, we find the DCT has little or no gain over computationally simpler approaches.

In our 2-D sub-band coder, we rely primarily on DPCM for sub-bands. We use a three-point causal predictor in each case, with the choice of pixels illustrated in Fig. 8 for the rectangular and quincunx- sampled bands. Our predicted value for the given pixel is

$$\hat{x}_0 = a_1 x_1 + a_2 x_2 + a_3 x_3.$$

The coefficients are chosen to yield a least mean-squared error prediction. Using common results from estimation theory, we can compute the coefficient vector $\mathbf{a}^T = [a_1, a_2, a_3]$ as

$$\mathbf{a} = R_{XX}^{-1} R_{X0}, \tag{9}$$

where R_{XX} is the autocorrelation matrix of the three pixel values given in Fig. 8, and R_{X0} is the cross-correlation vector between the value of x_0 and the three neighbors. We estimate

<div align="center">11</div>

R_{XX} and R_{X0} from several sample images from different sequences, then fix the prediction coefficients.

Such a predictive filter is very commonly thought of as useful for removing high correlation among neighboring pixels in a signal with much spectral energy concentrated at relativea low frequencies. In the present setting, it serves as an approximation to a whitening filter, which flattens the spectrum of the signal to be coded. This decorrelation lowers the entropy of each entry. The same principle holds for higher frequency sub-bands, which our predictor is designed to whiten, even if most of the energy is at the highest frequencies represented. Note that after decimation, the sub-band images are no longer bandpass at the lower resolution.

Following the predictive filters, we are left with prediction errors as the coefficients to be coded. Most research has indicated that for image representation, linear quantizers are more useful than the least mean-squared error Max quantizer, since infrequent, large errors detract more from image quality than quantitative measures indicate[15]. Each of our quantizers is linear, with increasingly large increasing quantizer step sizes as we move toward higher frequencies, and large dead zones about zero in the highest bands. This gradation has been quantified in terms of human perception by other investigators[15]. Low frequency errors due to quantization may contribute no more to the expected error magnitude per pixel than those in higher bands, but due to the large decimation ratios to which the lowest bands have been subjected, their errors translate into errors with extensive spatial correlations in the final reconstruction. As an extreme example, suppose the lowest sub-band consisted of only the zero-frequency value, a scalar. Any quantization error for this value results in the identical contribution to error over the entire image.

The spatial-domain coders described above accomplish significant reduction in entropy per sample, but more complete exploitation of these properties requires consideration of the one-dimensional distribution of prediction error in each band. The quantizers are therefore followed by Huffman coding for each band separately.

In typical natural imagery, highest frequency bands have low total energy, which is sparsely spatially distributed. Values of significant magnitude are often concentrated near sharp transitions of intensity. Many sub-band coders exploit this characteristic by recording some sort of information indicating locations of non-zero samples, in addition to the values of these samples[16, 14, 17]. For the highest band, if it is not deleted, we use as simple dead-zone quantizer, followed by run-length coding of strings of zeros.

12

3.3   Intra-Frame Coding Examples

The effects of the developed algorithms will be demonstrated on frames from three sequences. Each is given in CIF format, with resolution of 240 × 360 pixels. The originals are found in Figures 9-11. The "table tennis" sequence is the easiest of the three to compress, and "flowergarden" the most difficult. An example of the pyramidal decomposition of the "tennis" frame appears in Figures 12 and 13. Alternating sub-bands are rotated by 45 degrees for representation on the rectangular raster. Rather than decrease the physical size for visual presentation, we have used pixel replication with decreasing sample density at upper levels of the pyramid. In Figures 14-16, we have the same frames as above compressed by the pyramidal sub-band coder. Table 1 gives the parameters and statistics of each band and each image. In each of these, we have simply deleted the highest frequency band, HH of Fig. 1. This causes a small amount of aliasing, since the upper sub-band is not used to cancel it. The error is not easily noticed at normal viewing distance and frame rates.

The visual quality of these images is comparable to many coders, but the sub-band structure we've chosen has some important advantages, primarily in its concealment of relatively large quantitative error. The greatest part of the error energy results from the deletion of the HH band of Fig. 1, which causes the total luminance SNR in the results of Table 1 to be much lower than that of the coded bands. But this is error at frequencies to which we have little sensitivity. To illustrate the effect, we present the coded luminance portion of the "tennis" frame coded by JPEG standard in Fig. 17, and our sub-band coder in Fig. 18. The JPEG version is coded at a slightly higher bit rate, and has lower noise energy. But the JPEG artifacts, including blockiness in the background, and varying distortion of the table edge, are far more visually disturbing when viewed at normal video frame rates.
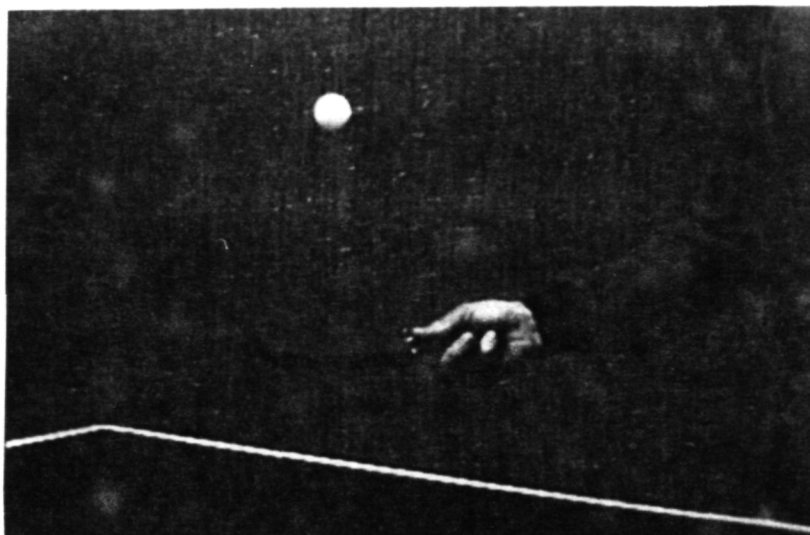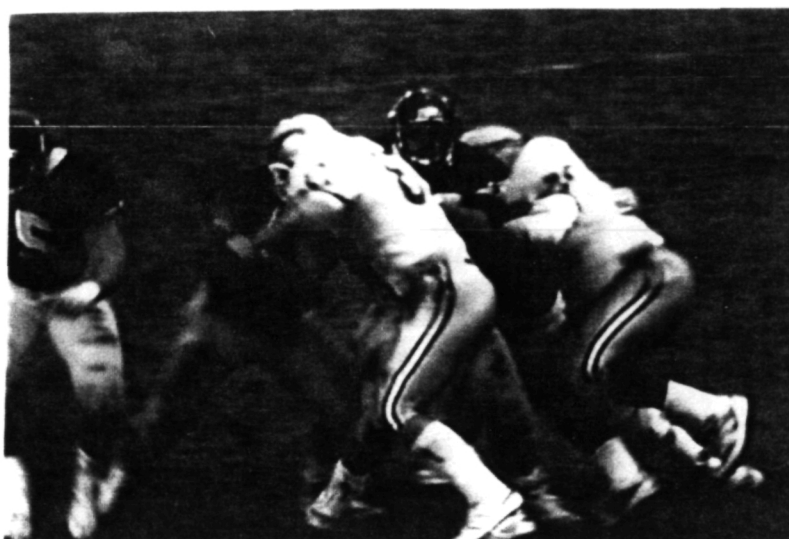
Figure 9: Original CIF image of frame 31 from "table tennis" sequence.

Figure 10: Original frame 0 from "football" sequence.

Figure 11: Original frame 0 from "flowergarden" sequence.

3.4 Edge-Adaptive Coding of High Frequency Sub-Bands

In addition to high frequency energy being somewhat spatially localized, its subjective perceptibility is dependent on the structure of which it is a part. High frequency loss is problematic primarily at sharp intensity edges, which are important visual cues to the human observer. One of the advantages of sub-band coding is the ease of adaptation to such specific temporal frequencies.

Transmission of very sparse spatial information such as single-band edge representation may be costly in terms of bit rate. It is imperative that any edge adaptation be made with minimal extra cost. Our approach is to add the highest frequency information only on edges which have perceptual importance in the image. These we hypothesize as those which have substantial low frequency in their structure, as well as high. We use simple one-dimensional edge detection along rows and columns of the rotated quincunx image, and code the highest band entries with a dead-zone quantizer only if they are part of the edge map. The edge detectors are low-pass filtered to eliminate non-macroscopic entities from consideration. The cost of specifying the *location* of non-zero samples in the highest band is potentially expensive. But we can perform the identical edge detection operation at both the coder and decoder, on the coded image, allowing the edge locations to be computed at
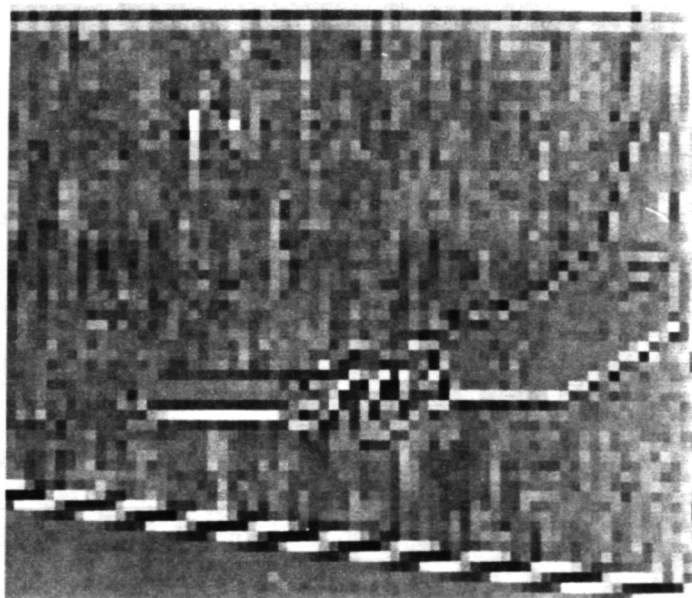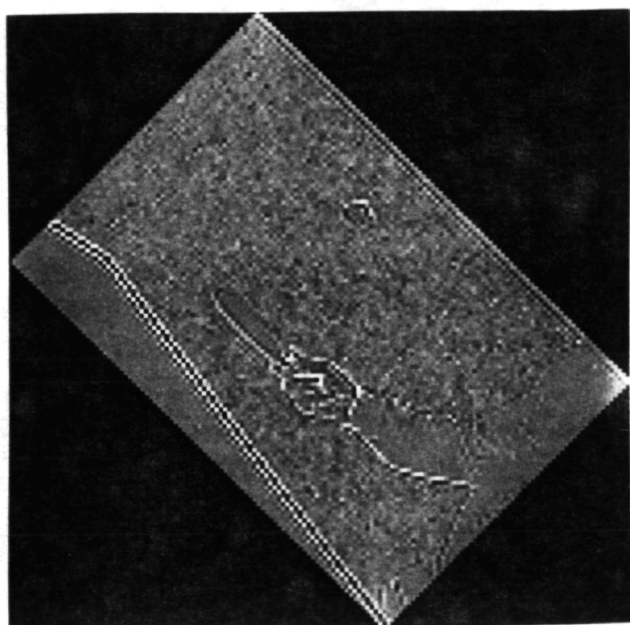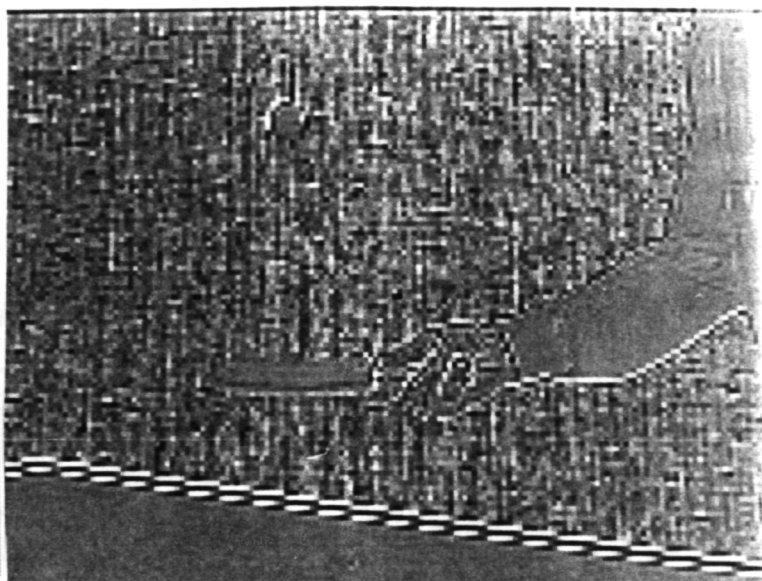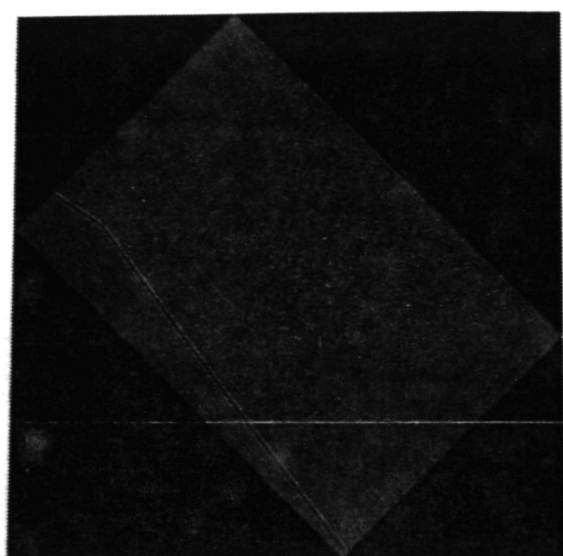
a	b
c	d

Figure 12: Pyramid decomposition of "tennis". All quincunx-sampled bands are rotated by 45 degrees. Images (a)-(d) correspond to progression of pyramid in Figure 7 up to $Y^{(4)}$. (b) and (d) are truncated at sides.
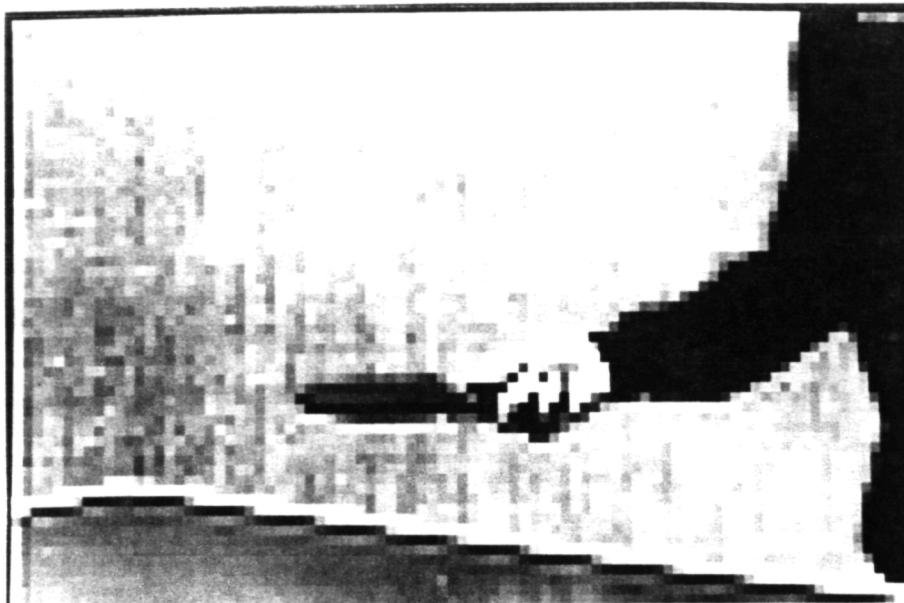
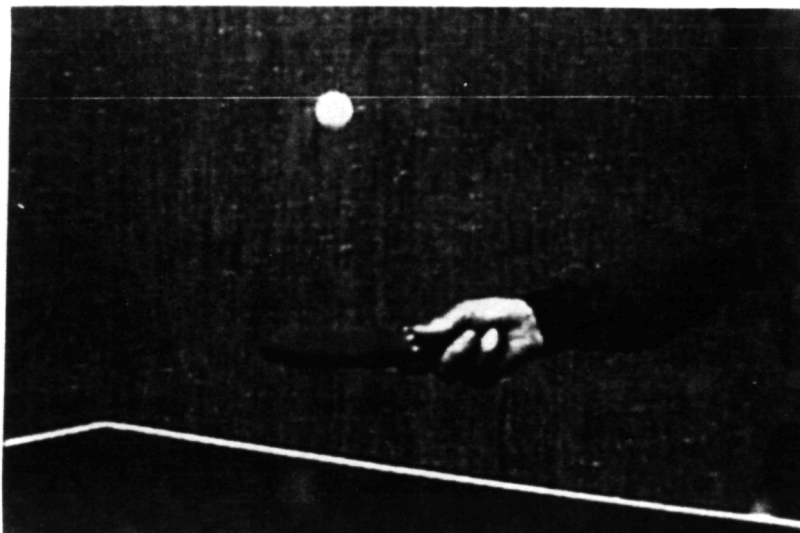Figure 13: Lowest sub-band ($Y^{(5)}$) from "tennis".

Figure 14: Frame 31 of "tennis" coded with 0.89 bpp.

17

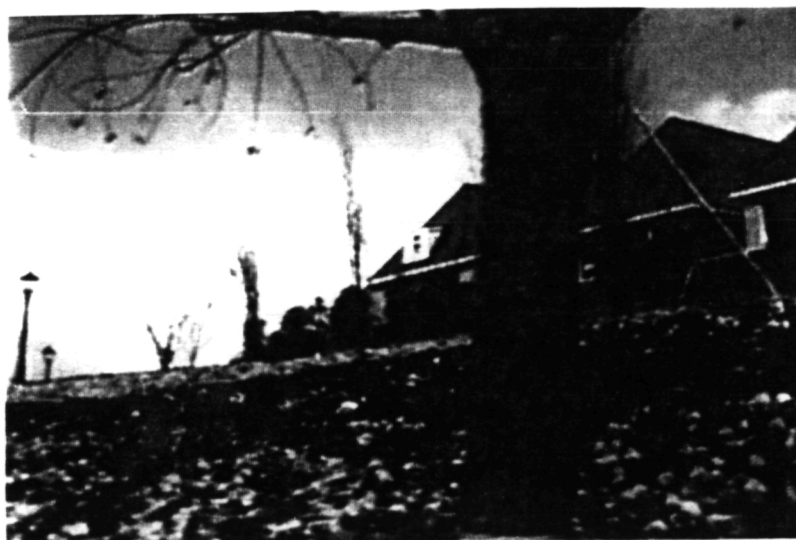Figure 15: Frame 0 of "football" coded with 1.15 bpp.

Figure 16: Frame 0 of "flowergarden" coded with 1.87 bpp.

Sequence	Sub-band	Deadzone	Quantizer Step	BPP	SNR(dB)
Table Tennis	$Y^{(1)}$	∞	0	0	0
	$Y^{(2)}$	15	4.0	0.13	34.1
	$Y^{(3)}$	0	5.0	0.26	45.1
	$Y^{(4)}$	0	3.5	0.17	48.1
	$Y^{(5)}$	0	3.0	0.21	49.3
	Total Lum.			0.77	29.0
	Cr			0.07	39.3
	Cb			0.05	40.0
Football	$Y^{(1)}$	∞	0	0.0	0
	$Y^{(2)}$	15	4.0	0.10	36.2
	$Y^{(3)}$	0	5.0	0.27	42.2
	$Y^{(4)}$	0	3.5	0.18	47.8
	$Y^{(5)}$	0	3.0	0.26	43.8
	Total Lum.			0.81	31.4
	Cr			0.15	40.0
	Cb			0.19	38.7
Flowergarden	$Y^{(1)}$	∞	0	0.0	0
	$Y^{(2)}$	15	4.0	0.45	33.0
	$Y^{(3)}$	0	5.0	0.40	39.2
	$Y^{(4)}$	0	3.5	0.23	46.8
	$Y^{(5)}$	0	3.0	0.32	46.2
	Total Lum.			1.40	23.8
	Cr			0.20	35.2
	Cb			0.27	34.0

Table 1: Bit rates and error for still-frame coding examples. $Y^{(1)}$, the highest non-rectangular frequency band, was omitted in all cases give here. This causes the relatively low SNR for the total luminance signal. "Cr" and "Cb" are red and blue color-difference representation of chrominance.
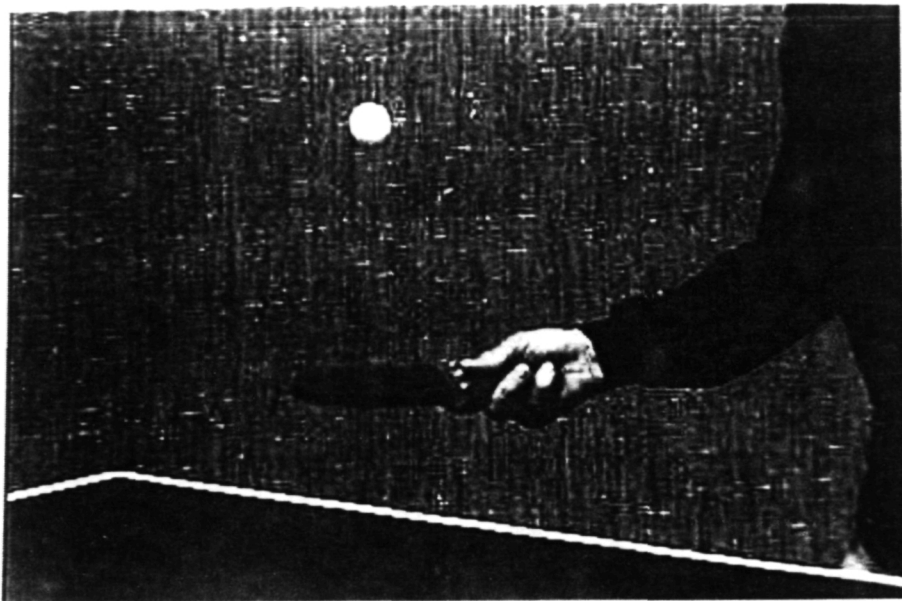
Figure 17: Luminance of Frame 31 of "tennis" coded with 0.81 bpp by JPEG algorithm. SNR is 30.8 dB.
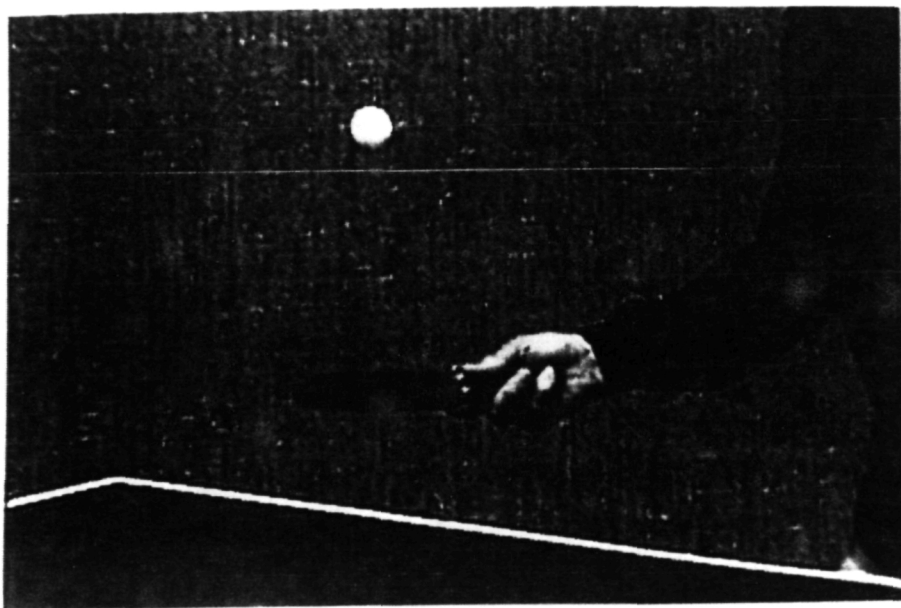

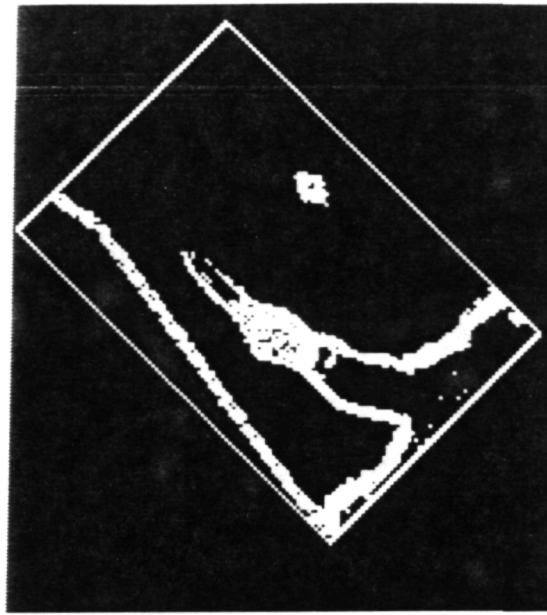Figure 18: Luminance coded with 0.77 bpp at 29.0 dB by sub-band algorithm.

Figure 19: Detected edges with highest band deleted, and quincunx sampling.

the receiver with no extra cost. The high-band non-zero samples are Huffman coded, and transmitted as a single string.

Figure 19 illustrates a typical detected set of edge pixels from "tennis." The edges represent the logical "OR" of the edges from the vertical and horizontal detection. This represents a single quincunx decimation by 2 for both the upper band, and the image in which edge detection is performed. Thus there is a 1-to-1 correspondence between pixels in the two images. As shown in Fig. 20. perceptible sharpness improvement results from the addition of high frequency edge information. The additional cost is about 0.06 bits per pixel, which may be a worthy investment, dependent on the specific application at hand. Fig. 21 includes the highest band, coded with dead-zone quantization and run-length coding, within the sub-band framework. Note that this improvement increases the bit rate by 0.11 bpp. Both methods effect essentially the same improvement in rendering edges, especially about the hands and table edge here. The edge adaptive approach, though, allows the lower bit rate by omitting high frequency in the background. At the range of bit rates we are considering, the omission may actually serve as an improvement, since the "speckled" non-zero pixels from the high frequency band produce a bothersome effect in the video. The trade-off between these two can be seen primarily in the background.
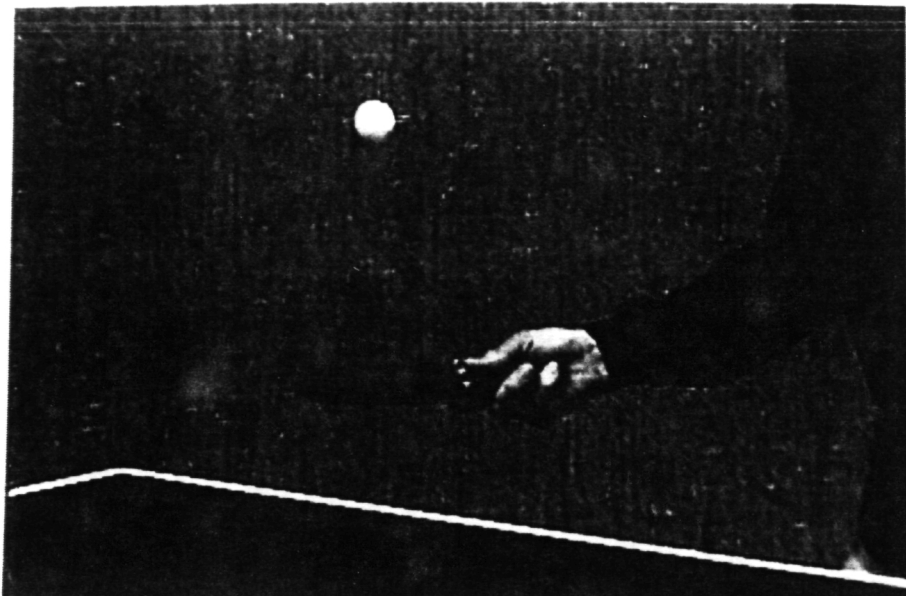
21

Figure 20: Luminance component of frame 31 of "tennis" coded with 0.83 bpp using edge-adaptive coding of highest frequency band.
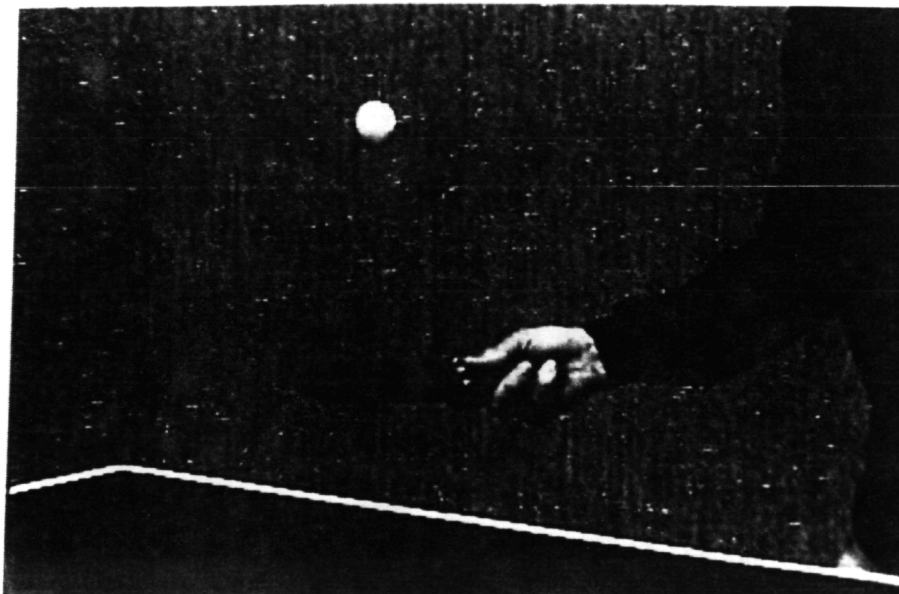

Figure 21: Luminance component of frame 31 with highest frequency sub-band coded by dead-zone quantization and run-length coder. Total rate = 0.88 bpp.

4   Motion Adaptive Sub-Band Video Sequence Coding

Video sequences of intelligible information include a high degree of temporal correlation. A large fraction of most sequences can be usefully described as a superposition of objects whose changes between adjacent frames in time over short periods consists of translational motion. This is the principle underlying motion estimation and compensation for sequence coding. But although motion compensation helps reduce bit rates, it adds a substantial computational burden and hardware cost to the system. In applications where bit rate constraints are not extremely stringent, and hardware costs are critical, simpler approaches may prove useful.

We have investigated image sequence coding with this type of application in mind, seeking a computationally simple coder to take advantage of temporal correlation. Research during the first period of the grant included the *estimation* of temporally active regions of a sequence, with the immediate goal of allowing conditional block replenishment. Here we generalize this idea to an adaptive coder, using the segmentation of an image in a sequence into temporally active and inactive areas to select a coding modality. In earlier versions of this work, before coding was employed, conditional replenishment was equivalent to the present case with the prediction error always quantized to zero. Because we have established with our previous work that we can effectively reduce bit rate in foreground/background settings to at most the product of the active percentage of a frame and the intraframe bit rate, we have of late studied only sequences in which the entire frame is in motion, due to panning and/or zooming of the camera.

Predictive coders obviously rely on a continuous "past" for their estimate of the current pixel value. The adaptivity of the coder on spatial blocks interrupts this pattern, complicating somewhat the process of coding most bands. Also, the high spatial frequencies are amplified by temporal differencing in moving areas. Due to these considerations, we apply a simpler sub-band technique, using the decomposition of Ansari, et al., with intra-frame compression of the LH band of Fig. 1, and temporal adaptivity on the LL band. Adaptivity consists of choosing the inter-frame difference signal, or the intraframe block for coding. Designation of each block's state is made strictly by the algorithm of [4], with the maximum-likelihood mode only. The LL band is coded by JPEG. Although we have experimented with several forms of the non-rectangular sub-band technique in this framework, we have seen little difference in bit rate improvements among distinct coders. In observing the first-order
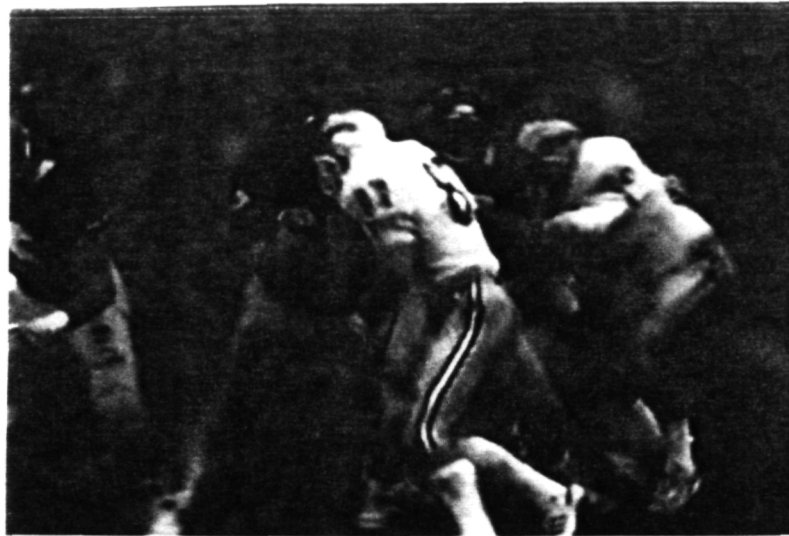
Figure 22: Frame 3 of "football" coded by adaptive inter/intraframe coder, using JPEG on either intraframe, or frame difference signal in LL band. Bit rate = 1.04 bpp.

effects of simple adaptivity on spatially disjoint blocks, the precise form of the coder on those blocks appears to be of less influence than the choice blocks for the different types of coding.

The results which we have achieved to this point are not conclusive concerning the adaptive 3-D coder. Bit rates are improved, but not yet enough to serve in many challenging problems. The compression gain in using our dynamics detection algorithm is approximately the same as is found by the much more costly technique of computing the energy in the interframe difference and intraframe (without zero-frequency element) image, and coding the smaller energy signal. Thus the detection method appears to work well.

A typical example is shown in Fig. 22, which includes much activity in the players' area, and is part of camera panning. Slightly over 50% of the frame was detected as active by the dynamics detection algorithm. This is due to the relatively low contrast change in the panning background. This frame's luminance is compressed to about 0.7 bpp, compared to 0.8 bpp using the hierarchical sub-band algorithm. While this bit rate represents a useful improvement, and attains approximately the original goal of 1 bpp for the sequence, other sequences, such as "flowergarden," are not likely to be sufficiently compressible.

5 Conclusion

Under the support of NASA Grant NAG3-1186, we have pursued several fundamental issues in the processing of image sequences to be presented as moving pictures. The final period has seen the development of a hierarchical coder which provides good performance with limited computational complexity, with the flexibility inherent in sub-band decomposition. This decomposition appears also to effectively hide quantitatively large errors very well, since it follows a frequency-domain pyramid more in concert with human visual sensitivities. The basic coder is established, but several aspects are worthy of further study. Although relatively long roll-off in our simple band-splitting filters has advantages, and causes no acute problems in coding of the entire pyramid structure, selective deletion of the upper bands for changes in resolution, transmission rates, etc., may dictate better cut-offs to avoid aliasing. Higher order filters should be studied for this case, both FIR and IIR.

Further study will also be allotted to the three-dimensional coding problem. One possible generalization is a more fully three-dimensional sub-band structure, such as has been applied to lower bit rate problems[14, 18]. It appears that completely general applicability of our approach will require the addition of motion compensation to the system. We are currently studying a new technique for computationally efficient motion estimation, which can run over an order of magnitude faster than the popular block-matching algorithm.

6 Appendix: Implementational Considerations for Filters

Processing speed is of utmost importance in video applications. The IIR filters we propose are especially simple, requiring only two multiplications per sample at each filtering. All filtering structures are separable into two orthogonal 1-D operations, functioning as one-dimensional filters applied to the diagonals of each sub-band image. This allows arbitrary degrees of parallelism, up to one processor per diagonal line of pixels across the image.

Implementation of the band-splitting filters in hardware introduces error in finite wordlength representation of coefficients and products. For simplest hardware realizations, we assume all computations are done in fixed-point format. For the first-order filter, the parameter a is critical in determining the propagation of quantization noise. We will denote the number of stages (levels in pyramid) as N. Three ranges of the value a are of particular interest:

- $|a| < \frac{\sqrt{2}-1}{2}$: Quantization error bound (Q) will increase with each stage, but reach a finite limit as $N \to \infty$.

- $|a| = \frac{\sqrt{2}-1}{2}$: $Q \approx 1.35Nq$, linear in the number of stages.

- $|a| > \frac{\sqrt{2}-1}{2}$: Q increases exponentially with N.

In the last case, additional bits may be needed for filtering operations.

Control of quantization error propagation is possible by increasing the number of bits used to represent each coefficient a_i in higher order filters. If $h(n)$ is the impulse response of the filter specified by $T(z)$, and q is the quantization step size for coefficients, the output error bound is

$$\frac{q}{4} \left(\sum_0^\infty |h(n)| \right)^2$$

for a single filter (highpass or lowpass). This assumes that a double-length accumulator with a rounding operation is used.

To a much greater extent than finite-impulse response (FIR) filters, recursive filters are susceptible to instability. Limit cycles are a particularly troublesome form in IIR filters. To guarantee the exclusion of limit cycles, we specify three ranges for the $\{a_i\}$, depending on the form of truncation used:

- Magnitude Truncation : $|a_i| < 1$
- Rounding : $|a_i| < 0.5$
- Two's Complement Truncation : $0 \le a_i < 1$

Thus for a first-order fixed-point implementation of the filters used in our work, a double length accumulator with rounding is optimal. Greater detail concerning limit cycle exclusion can be found in [19].

References

[1] P.H. Bauer, W. Qian, "A 3-D Nonlinear Recursive Digital Filter for Video Image Processing", *Proceedings of the 1991 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, Victoria B.C., Canada.

[2] C. Jones, "Sample Allocation for Data Compression in Video Sequences," *M.S. Thesis, University of Notre Dame, 1991.*

[3] C. Jones and K.D. Sauer, "A Bayesian Approach to Segmentation of Temporal Dynamics in Video Data," *Proc. SPIE Conf. Vis. Comm. and Im. Proc.*, Boston, MA, Nov. 10-13, 1991, pp. 522-533.

[4] K. Sauer and C. Jones, "Bayesian Block-Wise Segmentation of Interframe Differences in Video Sequences," submitted to *Computer Vision, Graphics, and Image Processing.*

[5] J.W. Woods and S.M. O'Neil, "Subband Coding of Images," *IEEE Trans. Acc., Speech, and Sig. Proc.*, vol. ASSP-34, no. 5, pp. 1278-1288, Oct. 1986.

[6] M. Vetterli, "Multi-dimensional Sub-band Coding: Some Theory and Algorithms," *Signal Processing*, vol. 6, pp. 97-112, April 1985.

[7] *JPEG Draft Technical Specification (Revision 5)*, Computer and Business Equipment Manufacturers Association, 311 First Street NW, Suite 500, Washington DC.

[8] R. Ansari, H.P. Gaggioni, and D.J. LeGall, "HDTV Coding Using a Nonrectangular Subband Decomposition," *Proc. SPIE Conf. on Vis. Comm. and Im. Proc.*, pp. 821-824, Cambridge, MA, Nov. 9-11, 1988.

[9] G. Schamel, "Motion Adaptive Four Channel HDTV Subband/DCT Coding," *Proc. SPIE Conf. on Visual Comm. and Im. Proc.*, Lausanne, Switzerland, Oct. 1-4, 1990, pp. 743-753.

[10] P.P. Vaidyanathan, "Perfect Reconstruction QMF Filter Banks for Two-dimensional Applications," *IEEE Trans. on Circ. and Sys.*, vol. CAS-34, pp. 976-978, May 1987.

[11] R. Ansari, A.M. Fernandez, and S.H. Lee, "High Definition Television Subband/Discrete Cosine Transform Coding Using Infinite-Duration Impulse Response Filter Banks: Coding Strategies," *Proc. SPIE Conf. on Vis. Comm. and Im. Proc.*, pp. 1291-1302, Philadelphia, PA, Nov. 8-10, 1989.

[12] T.A. Ramstad, "IIR Filterbank for Sub-band Coding of Images," *Proc. IEEE Int'l Symp. on Circ. and Sys.*, Espoo, Finland, June 1988, pp. 827-830.

[13] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *IEEE Sig. Proc. Magazine*, vol. 8, no. 4, pp. 14-38, Oct.. 1991.

[14] C.I. Podilchuk, N.S. Jayant, and P. Noll, "Sparse Codebooks for the Quantization of Non-Dominant Sub-Bands in Image Coding," *Proc. IEEE ICASSP 1990*, Albuquerque, NM, April 3-6, 1990, pp. 2101-2104.

[15] R. Safranek and J. Johnson, "A Perceptually Tuned Sub-band Image Coder with Image Dependent Quantization and Post-quantization Data Compression," *Proc. ICASSP 89*, Glasgow, Scotland, May 23-26, 1989, pp. 1945-1948.

[16] O. Johnson, O.V. Shentov, and S.K. Mitra, "A Technique for the Efficient Coding of the Upper Bands in Subband Coding of Images," *Proc. IEEE Int'l Conf. on Acoustics, Speech and Sig. Proc.*, Albuquerque, NM, April 3, 1990, pp. 2097-2100.

[17] D.J. LeGall, H.P. Gaggioni, and C.T. Chen, "Transmission of HDTV Signals Under 140 Mbits/sec Using Subband Decomposition and Discrete Transform Coding," *2nd Int'l Wkshp. on Sig. Proc. of HDTV*, Italy, March 1988.

[18] G. Karlsson and M. Vetterli, "Three-Dimensional Sub-Band Coding of Video," *Proc. ICASSP 1988*, New York, NY, pp. 1100-1103.

[19] P.H. Bauer and L.-J. LeClerc, "A Computer-Aided Test for the Absence of Limit Cycles in Fixed-Point Digital Filters," *IEEE Trans. Sig. Proc.*, vol. 39, no. 11, pp. 2400-2410, Nov. 1991.